

# Interaction-mediated surface state instability in disordered three-dimensional topological superconductors with spin $SU(2)$ symmetry

Matthew S. Foster<sup>1,\*</sup> and Emil A. Yuzbashyan<sup>2</sup>

<sup>1</sup>*Department of Physics and Astronomy, Rice University, Houston, Texas 77005, USA*

<sup>2</sup>*Center for Materials Theory, Department of Physics and Astronomy, Rutgers University, Piscataway, New Jersey 08854, USA*

(Dated: December 12, 2012)

We show that arbitrarily weak interparticle interactions destabilize the surface states of 3D topological superconductors with spin  $SU(2)$  invariance (symmetry class CI), in the presence of non-magnetic disorder. The conduit for the instability is disorder-induced wavefunction multifractality. We argue that time-reversal symmetry breaks spontaneously at the surface, so that topologically-protected states do not exist for this class. The interaction-stabilized surface phase is expected to exhibit ferromagnetic order, or to reside in an insulating plateau of the spin quantum Hall effect.

PACS numbers: 73.20.-r, 64.60.al, 05.30.Rt, 73.20.Fz

The existence of novel delocalized surface states is a key signature of 3D topological phases of matter [1–3]. These states envelop a fully-gapped, yet topologically “twisted” bulk and can display exceptional properties such as the quantized magnetoelectric effect and Majorana fermions [1]. A complete classification [2, 3] for (effectively) non-interacting particles has demonstrated that only five classes of topological phases and associated surface states arise in 3D.

An important development [2] has been the incorporation of disorder effects on 2D surface states. This is crucial because the terminating facets of a bulk 3D crystal inevitably host structural defects and impurities. The topologically nontrivial bulk is linked [2] to an effective low-energy surface theory of 2D Dirac fermions, perturbed by random impurity potentials [4, 5]. Each of the five classes of 3D topological phases is “protected” from the effects of time-reversal invariant (i.e., non-magnetic) disorder, in the sense that at least one surface Dirac wavefunction escapes Anderson localization [2, 5].

Unlike uniform plane waves, the extended 2D energy eigenstates enveloping a surface-disordered topological phase exhibit wild spatial amplitude fluctuations. These arise from quantum interference due to multiple impurity scattering, and manifest in the local density of states (LDOS). The pattern of LDOS fluctuations presents an intricate structure, characterized by an infinite set of scaling dimensions associated to interwoven fractal measures of the surface, a feature known as *multifractality* [5]. The evasion of localization in favor of multifractal scaling is rare in 2D, and is a signature of topological protection in the presence of disorder [6].

In this Letter, we show that topological protection can be undermined by interparticle interactions. In particular, we study the *combined* effects of multifractal LDOS fluctuations *and* interactions upon the surface Andreev bound states of 3D topological superconductors. Because the bulk condensate screens the long-ranged Coulomb force, surface quasiparticles interact only via

short-ranged potentials. In the clean limit, the vanishing density of states for the 2D Dirac surface band implies that weak short-ranged interactions are irrelevant, i.e. the surface states remain “protected.” However, it is known that disorder-induced LDOS multifractality can amplify interaction effects, such as pairing correlations near the superconductor-insulator transition [7]. With physics dominated by its surface, the complete picture of a 3D topological phase must incorporate *both* disorder-induced quantum interference and interactions [8].

Specifically, we demonstrate that arbitrarily weak interactions (consistent with bulk symmetries) destabilize the non-interacting surface states of 3D topological superconductors with spin  $SU(2)$  symmetry (class CI [2, 3]), in the presence of non-magnetic disorder. Multi-

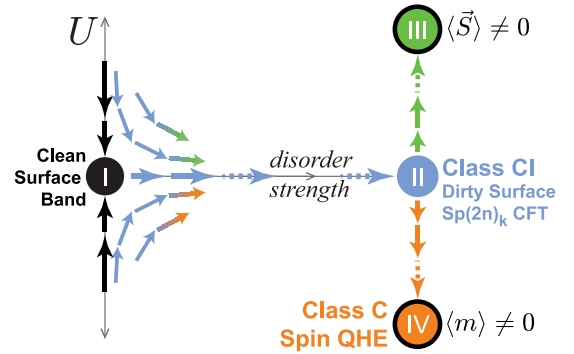


FIG. 1: Phase portrait sketch for the surface physics of a 3D time-reversal invariant, spin  $SU(2)$  symmetric topological superconductor. The vertical axis is the interaction strength  $U$  [Eq. (10)], while the horizontal axis measures non-magnetic disorder. Although the non-interacting system has a disorder-stabilized phase with delocalized (“protected”) surface states (II), it is destroyed by arbitrarily weak interactions [Eq. (11)]. Instead, at zero temperature, we expect that the surface exhibits broken spin symmetry [ $U > 0 \Rightarrow$  (III)], or the spin quantum Hall effect [ $U < 0 \Rightarrow$  (IV)]. In either scenario, interactions break time-reversal symmetry spontaneously.

fractal LDOS fluctuations enhance the interactions, facilitating the instability. We argue that time-reversal symmetry breaks spontaneously at the surface, so that “protected” surface states *do not exist* in this class. Depending upon the sign of the relevant interaction coupling  $U$  [see Fig. 1, Eqs. (10) and (11)], the surface should develop ferromagnetic order, or enter an insulating plateau state of the spin quantum Hall effect [9]. Our result provides impetus to identify a suitable material for the class CI bulk as an avenue to realize the spin quantum Hall phase. A similar analysis for the 3D topological superconductor class AIII will be published elsewhere [10].

Three of the five 3D topological symmetry classes can be realized as time-reversal invariant superconductors, distinguished by the degree of electronic spin conservation. In a 3D topological superconductor, Cooper pairing leads to a fully-gapped quasiparticle band in the bulk, associated to an integer-valued winding number  $\nu$  [2, 3]. The modulus  $|\nu|$  equals the number of flavors (or “valleys”) of 2D quasiparticle bands that appear at the sample surface, with energies that infiltrate the bulk gap. In the clean limit, the surface states exhibit a massless Dirac character at low energies; the Dirac point appears precisely at the bulk chemical potential (inside the gap) due to particle-hole symmetry.

In this paper, we study a universal low-energy model for the 2D surface states of a 3D class CI topological superconductor. In contrast to the spin-orbit-coupled  $\mathbb{Z}_2$  topological insulators, a CI superconductor has full spin  $SU(2)$  symmetry. The non-trivial topology arises through the entwining of orbital degrees of freedom, including non-(simple) s-wave pairing [11, 12]. For class CI  $\nu$  is even because Dirac surface bands appear in time-reversal conjugate pairs [2, 12, 13]. We consider the generic case with  $|\nu| = 2k$ ,  $k \in \{1, 2, 3, \dots\}$ . Neglecting interactions, the Hamiltonian for the surface theory is

$$H_D = \int d^2\mathbf{r} \psi^\dagger \{ -\hat{\sigma} \cdot [i\nabla - \mathbf{A}_i(\mathbf{r})\hat{\mathbf{t}}_\kappa^i] \} \psi. \quad (1)$$

The fermion field  $\psi$  is a complex Dirac spinor with pseudospin  $\sigma \in \{1, 2\}$  and valley  $v \in \{1, \dots, 2k\}$  indices, i.e.  $\psi \rightarrow \psi_{\sigma,v}$  when all indices are displayed. The pseudospin components  $\psi_{1,v}$  and  $\psi_{2,v}$  are linear combinations of the Nambu elements  $c_{\uparrow,v,\lambda}$  and  $c_{\downarrow,\bar{v},\lambda'}^\dagger$ . These annihilate (create) spin up (down) electrons in valley  $v$  ( $\bar{v}$ ). (Under time-reversal,  $v$  and  $\bar{v}$  interchange.) The indices  $\{\lambda, \lambda'\}$  label additional orbital (e.g. sublattice) degrees of freedom, whose precise interpretation depends upon bulk microscopics. A 3D class CI lattice model with  $\nu = \pm 2$  appeared in Ref. [12].

For a 3D topological superconductor, a key consequence of the non-trivial bulk is the special form that time-reversal symmetry adopts on the surface. If we write  $H_D \equiv \psi^\dagger \hat{h} \psi$ , with  $\hat{h}$  the single-particle Hamiltonian operator, then the usual time reversal symmetry for

spin-1/2 electrons in the bulk translates into the following *chiral* condition on the surface [2, 10, 12, 13]:

$$-\hat{\sigma}^3 \hat{h} \sigma^3 = \hat{h}. \quad (2)$$

This implies that any surface disorder that does not break time-reversal (including non-magnetic impurities) can manifest only as a random vector potential in the low-energy Dirac description. [Recall that  $\psi$  in Eq. (1) carries  $U(1)$  spin, rather than electric charge. In this language, vector potentials couple to time-reversal invariant spin and valley currents.] A homogeneous perturbation such as a chemical potential shift, or a time-reversal invariant pairing of the *surface* quasiparticles can be eliminated by a gauge transformation. Moreover, an energy gap (Dirac mass term) cannot appear at the surface of a topological superconductor unless time-reversal is broken. For class CI, non-magnetic disorder induces scattering between the  $2k$  valleys, in the form of the non-abelian valley vector potential  $\mathbf{A}_i(\mathbf{r})\hat{\mathbf{t}}_\kappa^i$  in Eq. (1). Here  $\hat{\mathbf{t}}_\kappa^i$  denotes a  $2k \times 2k$  generator of the group  $Sp(2k)$ . (The group is symplectic due to the spin symmetry [13].) In the absence of interactions, elastic scattering due to vector potential disorder begets delocalized, multifractal surface states, many properties of which can be computed exactly via conformal field theory (CFT) [4, 14, 15].

We first consider the effects of disorder upon the non-interacting surface states. Below we describe the physics and main idea of the CFT method. A technical summary can be found in Ref. [13], while a more comprehensive discussion will appear elsewhere [10]. The spatial character of the surface state wavefunctions (localized versus extended) can be ascertained via disorder-averaged moments of physical observables, such as the conductance or the local density of states (LDOS). To facilitate this, we replicate  $\psi_{\sigma,v} \rightarrow \psi_{\sigma,v,a}$ , where the replica index  $a \in \{1, \dots, n\}$ , and we are to take  $n \rightarrow 0$  at the end of the calculation [5]. Symmetry is the primary tool employed in the following, so we will rewrite Eq. (1) in a manifestly symmetric form. We decompose  $\psi$  and  $\psi^\dagger$  into “left”  $\mathcal{L}$  and “right”  $\mathcal{R}$  fields,

$$\begin{aligned} \{\mathcal{L}_{\uparrow,v,a}, \mathcal{L}_{\downarrow,v,a}\} &\equiv \left\{ \psi_{1,v,a}, \psi_{2,v',a}^\dagger (\hat{\kappa}^2)_{v',v} \right\}, \\ \{\mathcal{R}_{\uparrow,v,a}, \mathcal{R}_{\downarrow,v,a}\} &\equiv \left\{ \psi_{2,v,a}, \psi_{1,v',a}^\dagger (\hat{\kappa}^2)_{v',v} \right\}. \end{aligned} \quad (3)$$

Here and below, repeated indices are summed.  $\mathcal{L}_{s,v,a}$  denotes a  $4nk$ -component spinor; the index  $s$  ( $v$ ) transforms in the fundamental representation of the spin  $SU(2)$  [valley  $Sp(2k)$ ] symmetry. We also define

$$\mathbf{L} \equiv \mathcal{L}^\dagger i\hat{s}^2 \hat{\kappa}^2 \rightarrow \mathbf{L}_a^{s,v}, \quad \mathbf{R} \equiv \mathcal{R}^\dagger i\hat{s}^2 \hat{\kappa}^2 \rightarrow \mathbf{R}_a^{s,v}. \quad (4)$$

$\mathbf{L}_a^{s,v}$  and  $\mathbf{R}_a^{s,v}$  transform in the conjugate representations of the spin and valley symmetry groups;  $\hat{s}^2$  and  $\hat{\kappa}^2$  are spin and valley antisymmetric Pauli matrices [16].

Eq. (1) can be rewritten as  $H_D = H_0 + \delta H_D$ , where

$$H_0 = i \int d^2 \mathbf{r} [\mathbf{L} \bar{\partial} \mathcal{L} + \mathbf{R} \partial \mathcal{R}], \quad (5)$$

$$\delta H_D = \int d^2 \mathbf{r} [J_\kappa^i \bar{A}_i + \bar{J}_\kappa^i A_i]. \quad (6)$$

In Eq. (5), we have switched to complex spatial coordinates  $\{z, \bar{z}\} = x \pm iy$ ,  $\{\partial, \bar{\partial}\} \equiv \frac{1}{2}(\partial_x \mp i\partial_y)$ . The valley disorder appears in Eq. (6), where  $\{A_i, \bar{A}_i\} \equiv -i(A_x^i \mp iA_y^i)$ . This couples to the valley  $Sp(2k)$  current, which has the holomorphic component  $J_\kappa^i \equiv -(i/2)\mathbf{L} \hat{t}_\kappa^i \mathcal{L}$ .

Eq. (5) is manifestly invariant under chiral (independent left and right) spin  $SU(2)$ , valley  $Sp(2k)$ , and replica  $SO(n)$  transformations. The symmetry group enlarges to  $SO(4nk)$  if we include operations that mix all three index types. This free fermion theory is equivalent to the  $SO(4nk)_1$  Kac-Moody CFT [17]. The latter has a special property known as a conformal embedding rule [17–19], which gives a decomposition into a “product” of two other CFTs:  $Sp(2n)_k$ , associated to the (spin) $\times$ (replica) invariance of Eq. (5), and  $Sp(2k)_n$ , associated to the valley symmetry.

The delocalization physics of the non-interacting surface with Hamiltonian  $H_0 + \delta H_D$  is the same as in Refs. [14, 15], which dealt with 2D Dirac fermions coupled to a random  $SU(N)$  vector potential. Disorder is a relevant perturbation to the clean fermion theory [14]. Crucially, the impurity potential couples only to the valley current  $J_\kappa^i$  in Eq. (6). This leads to a “fractionalization” of the original  $SO(4nk)_1$  CFT: the valley  $Sp(2k)_n$  sector localizes, leaving behind the “critical” (delocalized) spin-replica  $Sp(2n)_k$  sector [14]. The latter is used to compute the scaling behavior disorder-averaged observables. Even in the absence of interactions, disorder localizes all surface states away from zero energy [20]; this is different from the case of a single Dirac fermion on the surface of a 3D topological insulator [21]. However, the localization length diverges upon approaching the chemical potential, so that the zero energy state at the Dirac point remains completely delocalized (“topologically protected”).

The disorder-induced spatial fluctuations of the LDOS  $\nu(\varepsilon, \mathbf{r})$  are encoded in the *multifractal spectrum*  $\tau(q)$  [5, 6]. The  $\tau(q)$  spectrum measures the sensitivity of extended wavefunctions to the sample boundary. A large  $L \times L$  area of the surface is finely partitioned into a grid of boxes of size  $a \ll L$ . One then defines the box probability  $\mu_n$  and inverse participation ratio  $\mathcal{P}_q$ ,

$$\mu_n(\varepsilon) \equiv \int_{\mathcal{A}_n} d^2 \mathbf{r} \nu(\varepsilon, \mathbf{r}), \quad \mathcal{P}_q(\varepsilon) \equiv \sum_n \left[ \frac{\mu_n(\varepsilon)}{\bar{\nu}} \right]^q, \quad (7)$$

where  $\mathcal{A}_n$  denotes the  $n^{\text{th}}$  box and  $\bar{\nu} = \sum_i \mu_i$  is the global DOS. When  $\varepsilon$  is tuned to a critical delocalization energy (such as a mobility edge),  $\mathcal{P}_q \sim (a/L)^{\tau(q)}$ , where the exponent  $\tau(q)$  is both self-averaging and universal [22]. The

multifractal spectrum thus provides a unique fingerprint for spatial fluctuations in a particular symmetry class. In the field-theoretic description, the  $q^{\text{th}}$  moment of the disorder-averaged LDOS ( $q \in \{1, 2, 3, \dots\}$ ) is associated to a particular composite operator  $\mathcal{O}_q$ , with scaling dimension  $\Delta_q$ . The set of such dimensions determines the multifractal spectrum via  $\tau(q) = 2(q-1) + \Delta_q - q\Delta_1$  [5, 6]. By contrast, localized states are insensitive to the sample boundary for sufficiently large  $L$  and have  $\tau(q) = 0$ .

For the class CI surface, we have identified the operators that represent disorder-averaged LDOS moments; these are a subset of the primary fields in the  $Sp(2n)_k$  CFT. As a result, we obtain the exact disorder-averaged multifractal spectrum at zero energy [10, 13],

$$\tau(q) = (q-1) \left[ 2 - \frac{q}{2(k+1)} \right]. \quad (8)$$

For  $k = 1$ , Eq. (8) agrees with previous calculations [15]; the form for general  $k$  is new. One of the main results of this paper, Eq. (8) proves that the non-interacting surface states at the bulk chemical potential remain delocalized, a consequence of the bulk topological order.

Now we turn to interparticle interactions. Robust surface states must be protected from the *combined* effects of both disorder and interactions. In a weakly-interacting fermion gas, the low-energy behavior of the density of states completely determines the importance of short-ranged interactions. The lowest-order (tree level) renormalization group (RG) equation for a generic four-fermion coupling  $U$  is [10]

$$d \ln U / dl = \Delta_1 - \Delta_2^{(U)} + \mathcal{O}(U), \quad (9)$$

where  $l$  denotes the log of the RG length scale such as the system size. In a *clean* 2D system,  $\Delta_2^{(U)} = 2\Delta_1$ , with  $\Delta_1$  the scaling dimension of the LDOS. For the clean Dirac surface band,  $\Delta_1 = 1$ , so that weak short-ranged interactions are strongly irrelevant. By contrast, a negative  $\Delta_1$  (due, e.g., to a van Hove singularity) would imply that  $U$  is relevant, signaling a potential instability. With impurities present, the exponents  $\Delta_1$  and  $\Delta_2^{(U)}$  denote scaling dimensions of the disorder-averaged LDOS and four-fermion interaction, respectively. The latter satisfies the lower bound  $\Delta_2^{(U)} \geq \Delta_2$  [10], where  $\Delta_2$  is the dimension of the second LDOS moment that determines  $\tau(2)$ . The crucial point is that  $\Delta_2$  is independent of, and *strictly less than*  $2\Delta_1$  for a multifractal delocalized state in a disordered system [23]. Impurity-mediated LDOS fluctuations can therefore amplify short-ranged interaction effects, by increasing the overlap of single particle wavefunctions in local regions. This is particularly relevant for an interaction  $U$  that saturates the bound  $\Delta_2^{(U)} = \Delta_2 < 2\Delta_1$ .

Physically, we expect that the important interactions include a spin exchange channel (because spin is a conserved hydrodynamic mode) and a Cooper pairing interaction (because disorder respects time-reversal). The

former is written  $-\vec{S} \cdot \vec{S}$ , where  $\vec{S}$  denotes the spin density. As discussed below Eq. (2), pairing of the surface quasiparticles does not open a gap unless time-reversal is simultaneously broken. The latter occurs when the Dirac mass operator  $m \equiv \psi^\dagger \hat{\sigma}^3 \psi$  develops an expectation value. This can be understood explicitly in the 3D CI topological superconductor lattice model of Ref. [12], which features real d-wave pairing in the bulk. In that model,  $m$  is interpreted as a sum of pairing operators:  $m \sim -ic_\uparrow^\dagger c_\downarrow^\dagger + ic_\downarrow c_\uparrow$ , where  $c_s$  annihilates a lattice electron. Crucially,  $m$  is odd under time-reversal, due to the factors of  $i$ . Thus, a non-zero expectation  $\langle m \rangle \neq 0$  would imply (“ $d + is$ ”) pairing of the surface quasiparticles, opening an energy gap and breaking time-reversal symmetry. The resulting state is an insulating plateau of the spin quantum Hall effect (see below). An attractive Cooper pairing interaction can be written as  $-m^2$ .

To keep the analysis general, we enumerate all four-fermion interactions that preserve the bulk symmetries [time-reversal invariance, spin  $SU(2)$ , and valley  $Sp(2k)$  symmetry]. This necessitates the incorporation of a third interaction channel  $J_S^\gamma \bar{J}_S^\gamma$ , where  $J_S^\gamma$  is the holomorphic spin current. The replicated interaction Hamiltonian for the CI surface is [13]

$$H_I = \sum_{a=1}^n \int d^2 \mathbf{r} \left[ U \left( m_a m_a - 4 \vec{S}_a \cdot \vec{S}_a \right) + V J_{S_a}^\gamma \bar{J}_{S_a}^\gamma + W \left( 3 m_a m_a + 4 \vec{S}_a \cdot \vec{S}_a - \frac{1}{k} J_{S_a}^\gamma \bar{J}_{S_a}^\gamma \right) \right]. \quad (10)$$

The interaction strengths  $\{U, V, W\}$  are defined so as to couple to RG eigenoperators, in the presence of disorder. In the minimal two valley realization ( $k = 1$ ), the  $W$ -channel interaction does not exist. For that case only,  $J_{S_a}^\gamma \bar{J}_{S_a}^\gamma = 3 m_a m_a + 4 \vec{S}_a \cdot \vec{S}_a$ .

Our task is to evaluate Eq. (9) in the disordered, non-interacting CI surface theory for the three interaction operators in Eq. (10) [24]. Using the  $Sp(2n)_k$  CFT, we have found that one particular operator controls the scaling of *both* the second LDOS moment and the interaction  $U$ , leading to  $\Delta_2^{(U)} = \Delta_2 = 0$ , while  $\Delta_1 = 1/2(k+1)$  [10, 13]. The main result of this paper follows,

$$\frac{dU}{dl} = \frac{U}{2(k+1)} + \mathcal{O}(U^2), \quad (11)$$

which implies that the interaction  $U$  in Eq. (10) grows at longer wavelengths, destabilizing the non-interacting, dirty surface, for any number of  $2k$  valleys. By contrast, the other interactions  $V$  and  $W$  remain irrelevant for any  $k$ , satisfying  $\frac{d \ln V}{dl} = -\frac{4k+3}{2(k+1)}$ ,  $\frac{d \ln W}{dl} = -\frac{3}{2(k+1)}$  [10, 13]. We conclude that while weak interactions are suppressed in the clean limit by the vanishing density of states at the Dirac point, surface disorder strongly renormalizes the interaction channel  $U$ , making it relevant.

Eq. (11) can be understood as an enhancement of interaction matrix elements in the eigenbasis of the disordered

theory: local accumulations of the DOS due to wavefunction multifractality induce stronger interactions between the surface quasiparticles. The amplification of the particular interaction channel  $U$  over the others signals the instability of the non-interacting surface to spontaneous time-reversal symmetry breaking. From Eq. (10), we anticipate (at least local) ferromagnetic order  $\langle \vec{S} \rangle \neq 0$  when  $U \rightarrow +\infty$ . Without time-reversal symmetry, the surface is not “topologically protected” [1–3], and we expect Anderson localization of all surface states [2, 5]. However, we cannot rule out an exotic metallic phase when spin symmetry is also broken [25]. By contrast,  $U \rightarrow -\infty$  should cause Cooper pairing of the surface quasiparticles. Treating the relevant interaction in mean field theory, one replaces  $m^2 \rightarrow 2\langle m \rangle \psi^\dagger \hat{\sigma}^3 \psi$  in Eq. (10). A non-zero Dirac mass opens an energy gap, producing an insulating surface. Time-reversal symmetry is broken because  $\langle m \rangle \neq 0$  implies surface pairing at a non-zero superfluid phase angle with respect to the bulk.

To lowest order in  $(1/k)$ , Eq. (11) agrees with a perturbative result [26] obtained using the non-linear sigma model [5, 13]. The calculations in Ref. [26] were performed in the context of a non-topological 2D system of gapless superconductor quasiparticles, subject to disorder and interactions with spin  $SU(2)$  symmetry and time-reversal invariance. The  $Sp(2n)_k$  CFT employed here has a sigma model representation with the same structure, but augmented with a Wess-Zumino-Witten (WZW) term [17]. In the  $k \gg 1$  limit, this model becomes weakly-coupled, and the WZW term can be ignored. The results of Ref. [26] therefore provide a non-trivial check of our analysis in the many-valley limit. In addition, at one loop in the sigma model calculation, RG flow equations beyond linear order in the interaction strengths can be obtained, because the sigma model treats interactions non-perturbatively via RPA and BCS-type summations. For the 2D class CI quasiparticle system, the sigma model generically predicts an instability of the “metallic” phase signaled by the divergence of the spin exchange or BCS pairing interaction strengths [13, 26]. This provides evidence for the absence of an interacting, time-reversal invariant fixed point.

The insulating state that occurs for  $\langle m \rangle \neq 0$  preserves spin  $SU(2)$  symmetry. This state resides in a plateau of the so-called spin quantum Hall effect [9], analogous to the “half-integer” quantum Hall phase at the surface of a 3D  $\mathbb{Z}_2$  topological insulator with broken time-reversal symmetry [1, 12]. The quantized spin Hall conductance [9] is  $\sigma_{xy}^s = \frac{1}{h} \left(\frac{\hbar}{2}\right)^2 p$ , with  $p = k \operatorname{sgn} \langle m \rangle$  if valley symmetry is unbroken *on average* (i.e., after disorder-averaging). If valley symmetry remains broken even after disorder-averaging, then  $p \in \{-k, -k+2, \dots, k-2, k\}$ ; see also Ref. [9]. Our results are summarized in Fig. 1.

In conclusion, we have demonstrated that interactions destabilize class CI disordered surface states in 3D. We



have argued that time-reversal symmetry breaks spontaneously, and that the CI topological superconductor surface enters into either a ferromagnetic or a spin quantum Hall phase. These are expected to be interaction-stabilized Anderson insulators. The other 3D topological superconductor classes AIII and DIII also admit WZW CFT descriptions [2]. The minimal surface state (single Dirac valley) realization for each of these is stable against disorder and short-ranged interactions [2, 10]. Results for class AIII with multiple valleys will appear elsewhere [10], while class DIII is an important topic for future work.

M.S.F. thanks Andreas Ludwig for helpful discussions. This work was supported by the NSF under Award No. DMR-0547769 and by the David and Lucille Packard Foundation.

- 
- \* Electronic address: matthew.foster@rice.edu
- [1] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010); X.-L. Qi and S.-C. Zhang, *ibid.* **83**, 1057 (2011).
  - [2] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).
  - [3] A. Kitaev, AIP Conf. Proc. No. 1134 (AIP, New York, 2009), p. 22.
  - [4] D. Bernard and A. LeClair, J. Phys. A **35**, 2555 (2002).
  - [5] F. Evers and A. D. Mirlin, Rev. Mod. Phys. **80**, 1355 (2008).
  - [6] M. S. Foster, Phys. Rev. B **85**, 085122 (2012).
  - [7] T. I. Baturina *et al.*, Physica C **468**, 316 (2008); M. V. Feigel'man, L. B. Ioffe, V. E. Kravtsov, and E. A. Yuzbashyan, Phys. Rev. Lett. **98** 027001 (2007).
  - [8] P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, Phys. Rev. Lett. **105**, 036803 (2010).
  - [9] I. A. Gruzberg, A. W. W. Ludwig, and N. Read, Phys. Rev. Lett. **82**, 4524 (1999); T. Senthil, J. B. Marston, and M. P. A. Fisher, Phys. Rev. B **60**, 4245 (1999).
  - [10] M. S. Foster, unpublished.
  - [11] N. Read and D. Green, Phys. Rev. B **61**, 10267 (2000).
  - [12] A. P. Schnyder, S. Ryu, and A. W. W. Ludwig, Phys. Rev. Lett. **102**, 196804 (2009).
  - [13] See supplementary information.
  - [14] A. A. Nersisyan, A. Tsvelik, and F. Wenger, Phys. Rev. Lett. **72**, 2628 (1994).
  - [15] C. Mudry, C. Chamon, and X.-G. Wen, Nucl. Phys. B **466**, 383 (1996); J. S. Caux, I. I. Kogan, A. M. Tsvelik, *ibid.* **466**, 444 (1996).
  - [16] In Eq. (4), “T” is the matrix transpose, while  $\hat{s}^2$  ( $\hat{\kappa}^2$ ) denotes the  $2 \times 2$  ( $2k \times 2k$ ) antisymmetric invariant tensor for the spin  $SU(2)$  [valley  $Sp(2k)$ ].
  - [17] P. Di Francesco, P. Mathieu, and D. Sénéchal, *Conformal Field Theory* (Springer-Verlag, New York, 1996).
  - [18] J. Fuchs, *Affine Lie Algebras and Quantum Groups* (Cambridge University Press, Cambridge, England, 1992).
  - [19] The relevant conformal embedding is  $SO(4nk)_1 \supset Sp(2n)_k \oplus Sp(2k)_n$  [18].
  - [20] At non-zero energy, all states are localized. Energy breaks the particle-hole symmetry, altering the class to AI. This has no delocalized phase in 2D [2, 5].

- [21] J. H. Bardarson, J. Tworzydło, P.W. Brouwer, and C.W. J. Beenakker, Phys. Rev. Lett. **99**, 106801 (2007); S. Ryu, C. Mudry, H. Obuse, and A. Furusaki, *ibid.* **99**, 116601 (2007); K. Nomura, M. Koshino, and S. Ryu, *ibid.* **99**, 146806 (2007); P. M. Ostrovsky, I. V. Gornyi, and A. D. Mirlin, *ibid.* **98**, 256801 (2007).
- [22] C. C. Chamon, C. Mudry, and X.-G. Wen, Phys. Rev. Lett. **77**, 4194 (1996).
- [23] B. Duplantier and A. W. W. Ludwig, Phys. Rev. Lett. **66**, 247 (1991).
- [24] D.-H. Lee and Z. Wang, Phys. Rev. Lett. **76**, 4014 (1996).
- [25] J. T. Chalker *et al.*, Phys. Rev. B **65**, 012506 (2001).
- [26] L. Dell’Anna, Nucl. Phys. B **758**, 255 (2006).

## SUPPLEMENTARY INFORMATION

In this supplement, we provide the minimal details necessary to reproduce the results in the main text; a comprehensive discussion and analysis of class CI and AIII topological superconductors will appear elsewhere [1]. We first summarize the symmetry structure of the class CI surface state theory. Then we provide the derivation of the multifractal spectrum and interaction scaling dimensions. We close with a review of the many-valley limit and the relation to the non-linear sigma model.

### SYMMETRIES OF THE SURFACE STATE THEORY

In Eq. (1) of the main text, the Dirac fermion  $\psi_{\sigma,v}(\mathbf{r})$  is a surface-confined projection [2–4] of a bulk field  $\Psi_{\sigma,v,\tau}(\mathbf{r}, z)$  that lowers the spin angular momentum by one. Relative to the surface, the bulk fermion carries an additional index  $\tau \in \{1, 2\}$ . The four  $\sigma \otimes \tau$  components of  $\Psi_v$  are linear combinations of electron operators  $c_{\uparrow,v,\lambda}$  and  $c_{\downarrow,\bar{v},\lambda'}^\dagger$ , where  $(v, \bar{v})$  denote a pair of valleys related by time-reversal, and  $\{\lambda, \lambda'\}$  run over additional orbital (e.g. sublattice) labels [1, 4].

In the bulk, a physical time-reversal operation sends  $i \rightarrow -i$  (complex conjugation) and

$$c_{\uparrow,v} \rightarrow -\hat{M} c_{\downarrow,\bar{v}}, \quad c_{\downarrow,v} \rightarrow \hat{M} c_{\uparrow,\bar{v}}.$$

In this equation,  $\hat{M} \rightarrow M_{\lambda,\lambda'}$  is a symmetric, unitary matrix in orbital labels. Time-reversal induces the transformation  $\Psi \rightarrow -i\hat{\sigma}^3(2\hat{\mathbf{P}} - \hat{1})\Psi^\dagger$ , where  $\hat{\mathbf{P}} = (\hat{1} + \hat{n} \cdot \hat{\boldsymbol{\tau}})/2$  projects onto a certain  $\tau$ -spin direction  $\hat{n}$ . (The latter is basis-dependent and determined by microscopics [1].) At the surface, the  $\tau$ -spin becomes “locked,” with  $\psi(\mathbf{r}) \sim \hat{\mathbf{P}}\Psi(\mathbf{r}, z=0)$  [3, 4]. This type of projection always occurs at the surface of a  $d$ -dimensional topological insulator or superconductor, producing an anomalous state that is “half” of a normal  $(d-1)$ -dimensional system.

For the surface theory in Eq. (1), time-reversal therefore appears as the antiunitary transformation [3–5]

$$\psi \rightarrow -i\hat{\sigma}^3 [\psi^\dagger]^\top, \quad i \rightarrow -i. \quad (\text{S1})$$

Spin  $SU(2)$  symmetry requires invariance under  $U(1)$   $\psi \rightarrow e^{i\theta}\psi$  and particle-hole  $\psi \rightarrow i\hat{\sigma}^1\hat{\kappa}^2 [\psi^\dagger]^\top$  transformations; the latter encodes a  $\pi$  rotation by  $S^x$ . Imposing these upon Eq. (1) restricts the disorder to the  $Sp(2k)$  valley vector potential  $\mathbf{A}_i \hat{\mathbf{t}}_\kappa^i$ , where  $-\hat{\kappa}^2 [\hat{\mathbf{t}}_\kappa^i]^\top \hat{\kappa}^2 = \hat{\mathbf{t}}_\kappa^i$ . In these equations,  $\hat{\kappa}^2$  denotes the valley  $Sp(2k)$  invariant tensor (antisymmetric  $2k \times 2k$  block Pauli matrix).

### CONFORMAL EMBEDDING

The disordered, non-interacting class CI surface state theory is described by the  $Sp(2n)_k$  Kac-Moody CFT. To see this, we note that the free fermion Hamiltonian in Eq. (5) of the main text is equivalent to the  $SO(4nk)_1$  current algebra [6]. (Since we are discussing non-interacting fermions at this stage, we can trade the 2D Hamiltonian for a 2+0-D Grassmann field action.) The conformal embedding  $SO(4nk)_1 \supset Sp(2n)_k \oplus Sp(2k)_n$  [7] implies that the associated stress tensor  $T(z)$  has the Sugawara decomposition [6]

$$T(z) = \frac{1}{2(n+k+1)} \left[ : J_\kappa^i J_\kappa^i : (z) + : J_{SR}^\alpha J_{SR}^\alpha : (z) \right], \quad (\text{S2})$$

where

$$J_\kappa^i \equiv \frac{i}{2} \sum_a J_{\kappa a}^i, \quad J_{\kappa a}^i \equiv -\mathbf{L}_a \hat{\mathbf{t}}_\kappa^i \mathcal{L}_a, \quad J_{SR}^\alpha \equiv -\frac{i}{2} \mathbf{L} \hat{\mathbf{t}}_{SR}^\alpha \mathcal{L}. \quad (\text{S3})$$

The replica-summed valley current  $J_\kappa^i(z)$  satisfies the  $Sp(2k)_n$  algebra. In Eqs. (S2) and (S3),  $J_{SR}^\alpha$  generates  $Sp(2n)$  spin  $\times$  replica space transformations;  $(\hat{\mathbf{t}}_{SR}^\alpha)_{s'aa'}$  is a suitable  $2n \times 2n$  matrix. The current  $J_{SR}^\alpha(z)$  satisfies the  $Sp(2n)_k$  algebra.

The impurity potential  $\{A_i, \bar{A}_i\}$  in Eq. (6) of the text couples only to the valley Kac-Moody current  $\{J_\kappa^i, \bar{J}_\kappa^i\}$ . As the disorder renormalizes towards strong coupling, it localizes the valley  $Sp(2k)_n$  sector. To obtain Eqs. (8) and (11), we express the associated operators in terms of  $Sp(2n)_k$  primary fields and descendants. The most relevant components govern the leading scaling behavior.

## MULTIFRACTAL SPECTRUM

We consider first the  $\tau(q)$  spectrum at level  $k = 1$ . Primary fields are labeled by the  $Sp(2n)$  fundamental weights  $\{\Lambda_q\}$ ,  $q \in \{1, \dots, n\}$  [6]. Weight  $\Lambda_q$  corresponds to a rank  $q$ , fully antisymmetric tensor  $\Omega_{[i_1 i_2 \dots i_q]}^{(q)}$ , satisfying the traceless condition  $(\hat{s}^2)^{ij} \Omega_{[ij \dots i_q]} = 0$ . Here,  $i = \{s, a\}$  is a product of spin  $s$  and replica  $a$  indices;  $(\hat{s}^2)^{ij}$  is the  $Sp(2n)$  invariant tensor. In the holomorphic sector, the  $q^{\text{th}}$  LDOS moment must correspond to a tensor with  $q$  distinct replica indices. We therefore associate

$$\begin{aligned} & \overline{(\psi_1^\dagger \psi_1)(\psi_2^\dagger \psi_2) \times \dots \times (\psi_q^\dagger \psi_q)(\mathbf{r})} \\ & \Leftrightarrow \Omega_{[\{s_1, 1\}\{s_2, 2\} \dots \{s_q, q\}]}^{(q)}(z) \bar{\Omega}_{[\{s'_1, 1\}\{s'_2, 2\} \dots \{s'_q, q\}]}^{(q)}(\bar{z}) \\ & \times (i\hat{s}^2 \hat{s}^3)^{s_1, s'_1} (i\hat{s}^2 \hat{s}^3)^{s_2, s'_2} \times \dots \times (i\hat{s}^2 \hat{s}^3)^{s_q, s'_q}. \end{aligned} \quad (\text{S4})$$

The left-hand side is the disorder-averaged  $q^{\text{th}}$  moment. The right-hand side is a diagonal primary field labeled by  $\Lambda_q$ . Here we have written each  $Sp(2n)$  index as the spin-replica product  $\{s, a\}$ ;  $\psi^\dagger \psi \rightarrow S^3$  is the *spin-projected* LDOS [8] for the surface state. We find that the most relevant contribution to the  $q^{\text{th}}$  LDOS moment is associated to the same  $Sp(2n)$  representation  $\Lambda_q$ , for any  $k \geq 1$ . The scaling dimension of the  $q^{\text{th}}$  moment is then given by [6]

$$\Delta_q = (2q - q^2)/2(k + 1) \quad (\text{S5})$$

in the replica limit [9]  $n \rightarrow 0$ . Eq. (8) follows.

## INTERACTION DIMENSIONS

The flow equations for each of the three interaction couplings  $U$ ,  $V$ , and  $W$  in Eq. (10) of the main text appear as in Eq. (9), with  $\Delta_1$  given as above and  $\Delta_2^{(U, V, W)}$  the scaling dimension of the associated field. We rewrite Eq. (10) exploiting  $Sp(2n)$  [10] and  $SU(2)$  Fierz identities:

$$H_I = \int d^2\mathbf{r} \left[ 2U J_{\kappa a}^i \bar{J}_{\kappa a}^i + V J_{S a}^\gamma \bar{J}_{S a}^\gamma + 2W I_{v' a}^{v\gamma} \bar{I}_{v a}^{v'\gamma} \right], \quad (\text{S6})$$

where  $J_{S a}^\gamma \equiv -\mathbf{L}_a \hat{s}^\gamma \mathcal{L}_a$  denotes a replica-resolved spin current (no sum on  $a$  is implied). Here  $\hat{s}^\gamma$  denotes a spin space Pauli matrix. The holomorphic half of the last term in Eq. (S6) is  $I_{v' a}^{v\gamma} \equiv \mathbf{L}_a^v \hat{s}^\gamma \mathcal{L}_{v' a} + \delta_{v' a}^v \frac{1}{2k} J_{S a}^\gamma$ . If we assume that the disorder-averaged theory is invariant under both spin  $SU(2)$  and valley  $Sp(2k)$  transformations, then the three channels  $U$ ,  $V$ , and  $W$  in Eq. (S6) exhaust the possibilities for four-fermion interactions.

As  $V$  couples to a KM current-current perturbation,  $\Delta_2^{(V)} = 2$  [6]. The  $U$  interaction involves the valley current  $J_{\kappa a}^i$ . This is *not* a valley KM current, but rather a product of  $Sp(2n)_k$  and  $Sp(2k)_n$  primary fields. [Replica-resolved components cannot be extracted from the KM current  $J_\kappa^i$  in Eq. (S3).] In the  $Sp(2n)_k$  theory,  $J_{\kappa a}^i$  corresponds to a second rank tensor with equal replica indices, antisymmetrized over spin (to obtain a singlet). The only choice is  $J_{\kappa a}^i(z) \Leftrightarrow \Omega_{[\{s, a\}\{s', a\}]}^{(2)}(z)$ , i.e. the same representation  $\Lambda_2$  that determines the scaling of the second LDOS moment [Eqs. (S4) and (S5) with  $q = 2$ ]. As a result,  $\Delta_2^{(U)} = \Delta_2 = 0$ . Finally, the  $W$  interaction operator  $I_{v' a}^{v\gamma}$  is a second rank tensor field with equal replica indices, symmetric in spin. The only choice has weight  $2\Lambda_1$ , leading to  $\Delta_2^{(W)} = 2/(k + 1)$  [6]. Via Eq. (9), we obtain Eq. (11) and the equations for  $V$  and  $W$  quoted in the text.

## MANY VALLEY (LARGE $k$ ) LIMIT: REVIEW OF THE CLASS CI FINKEL'STEIN NL $\sigma$ M RESULTS

In Ref. [11], the effects of interactions upon gapless quasiparticles in disordered, non-topological 2D superconductors were considered using the Finkel'stein non-linear sigma model framework [12]. In particular, the author studied symmetry class CI, appropriate to a spin singlet superconductor possessing spin  $SU(2)$  symmetry and time-reversal invariance in every realization of the disorder. The 2D topological surface state  $Sp(2n)_k$  CFT studied in the present paper also possesses a sigma model description with the same structure, but augmented by a WZW term [3, 5, 9, 13]. In the limit of many valleys  $k \gg 1$ , the WZW term can be ignored in the first approximation, yielding a “metallic phase” with a large spin conductance proportional to  $k$ , independent of the disorder [9]. The advantage of working directly

in the sigma model framework is that interactions are treated to all orders, via RPA and BCS-type resummations. The disadvantage is that the sigma model becomes strongly coupled for small  $k \in \{1, 2, 3, \dots\}$ , so that the WZW term cannot be ignored and perturbation theory becomes unreliable.

The only interactions that appear in the CI sigma model treated in Ref. [11] are short-ranged, and include a spin triplet exchange coupling  $\gamma_t$  and a Cooper pairing interaction  $\gamma_c$ . Here, we define  $\gamma_{t,c} > 0$  for repulsive interactions in each channel; these are dimensionless couplings in the limit of vanishing disorder strength. Other (e.g. electric charge density-density) interaction channels do not couple to conserved hydrodynamic diffusion modes, and are strongly irrelevant in the sigma model framework.

The one-loop RG equations for  $\gamma_{t,c}$  are [11]

$$\begin{aligned}\frac{d\gamma_t}{dl} &= -\frac{\lambda}{2}\gamma_c(1-\gamma_t)(1-2\gamma_t), \\ \frac{d\gamma_c}{dl} &= \frac{\lambda}{2}\{-3\gamma_t-2\gamma_c+3\gamma_c[\log(1-\gamma_t)+\gamma_t]\}-\gamma_c^2.\end{aligned}\tag{S7}$$

In these equations,  $\lambda$  is proportional to the dimensionless inverse spin conductance. For weak interactions, these equations can be linearized in the coupling strengths. The result is

$$\frac{d\ln\gamma_r}{dl} = \frac{\lambda}{2}, \quad \frac{d\ln\gamma_i}{dl} = -\frac{3\lambda}{2},\tag{S8}$$

where  $\gamma_r \equiv \gamma_c - 3\gamma_t$ , and  $\gamma_i \equiv \gamma_c + \gamma_t$ . Since  $\lambda \propto 1/k$  [6, 9], we find that Eq. (S8) matches the RG equations for the relevant ( $U \leftrightarrow \gamma_r$ ) and irrelevant ( $W \leftrightarrow \gamma_i$ ) coupling strengths in the WZW surface state theory studied in the present paper, Eq. (11) and the text following, valid to lowest order in  $1/k$ . The relevant sigma model coupling  $\gamma_r$  is indeed the difference of repulsive singlet pairing and triplet exchange interactions, just as we found in the WZW model.

Integrating the full flow Eqs. (S7) numerically, one sees that the metallic phase is generically destroyed by one of two instabilities: either the triplet exchange interaction flows to minus infinity  $\gamma_t \rightarrow -\infty$  (suggesting Stoner ferromagnetism), or the Cooper pairing strength flows to minus infinity  $\gamma_c \rightarrow -\infty$ , indicating a residual pairing instability for the gapless quasiparticles. We do not discuss here the back-reaction of the interactions upon the spin conductance [11], since the lowest order quantum corrections are modified by the WZW term.

---

\* Electronic address: matthew.foster@rice.edu

- [1] M. S. Foster, unpublished.
- [2] M. Z. Hasan and C. L. Kane, Rev. Mod. Phys. **82**, 3045 (2010); X.-L. Qi and S.-C. Zhang, *ibid.* **83**, 1057 (2011).
- [3] A. P. Schnyder, S. Ryu, A. Furusaki, and A. W. W. Ludwig, Phys. Rev. B **78**, 195125 (2008).
- [4] A. P. Schnyder, S. Ryu, and A. W. W. Ludwig, Phys. Rev. Lett. **102**, 196804 (2009).
- [5] D. Bernard and A. LeClair, J. Phys. A **35**, 2555 (2002).
- [6] P. Di Francesco, P. Mathieu, and D. Sénéchal, *Conformal Field Theory* (Springer-Verlag, New York, 1996).
- [7] J. Fuchs, *Affine Lie Algebras and Quantum Groups* (Cambridge University Press, Cambridge, England, 1992).
- [8] M. S. Foster, Phys. Rev. B **85**, 085122 (2012).
- [9] F. Evers and A. D. Mirlin, Rev. Mod. Phys. **80**, 1355 (2008).
- [10] For suitably normalized  $\{\hat{\mathbf{t}}_{\kappa J}^i\}$ ,  $\sum_i (\hat{\mathbf{t}}_{\kappa}^i \hat{\kappa}^2)^{v_1, v_2} (\hat{\kappa}^2 \hat{\mathbf{t}}_{\kappa}^i)^{v_3, v_4} = \frac{1}{2} [\delta^{v_1, v_4} \delta^{v_3, v_2} + \delta^{v_1, v_3} \delta^{v_4, v_2}]$ .
- [11] L. Dell'Anna, Nucl. Phys. B **758**, 255 (2006).
- [12] A. M. Finkel'stein, Zh. Eksp. Teor. Fiz. **84**, 168 (1983), Sov. Phys. JETP **57**, 97 (1983); for a review, see e.g. D. Belitz and T. R. Kirkpatrick, Rev. Mod. Phys. **66**, 261 (1994).
- [13] A. A. Nersisyan, A. Tselik, and F. Wenger, Phys. Rev. Lett. **72**, 2628 (1994).